Urban monitoring using NetKDE and VGI

Network based kernel density estimation on volunteered geographic information applied to Baghdad, Iraq

Nicolas LACHANCE-BERNARD[1]; Timothée PRODUIT[1]; Loic GASSER[1]

[1]Laboratory of Geographic Informations Systems, Ecole Polytechnique Fédérale de Lausanne
EPFL ENAC IIE LASIG, Bâtiment GC, Station 18, CH-1015 Lausanne, Switzerland
+41 (21) 693 57 87, Nicolas.Lachance-Bernard@epfl.ch (correspondent author),
+41 (21) 693 57 83, Timothee.Produit@epfl.ch, Loic.Gasser@epfl.ch

Keywords: Network based kernel density estimation (NetKDE), Spatio-temporal analysis, Fuzzy-set map comparison, OpenStreetMap, Volunteered geographic information (VGI)

Introduction

This paper presents a methodology for urban monitoring using volunteered geographic information (VGI) and journalism data Iraq war logs with network based kernel density estimation (NetKDE). It investigates, using spatio-temporal analysis, the evolution of urban events in Baghdad between 2004 and 2009. The extracted street network is based on the data distributed by OpenStreetMap (OSM). A total of 21,876 logged events, 66,648 network segments, 22,644 gridpoints (200m resolution grid) and 362,304 gridpoints (50m resolution grid) are used for the analysis. The methodology combines and adapts these VGI data and is mainly based on open source and/or publicly available software. It handles very large datasets with multi-scale, multi-resolution and temporal perspectives. Fuzzy-set map comparison (FMC) is used to identify level of changes between each period of time. The methodology is already used in other fields of research being biology, urban planning, criminology or economic development. It should help stakeholders in respective domain to analyze the evolution of network constrained events in multiple contexts.
This paper is divided in three parts. Firstly, conceptual background of VGI, NetKDE and FMC is presented. Secondly, the methodology is illustrated using data Iraq war logs, OSM data and grids with two different resolutions. Thirdly, spatio-temporal analysis results are presented and discussed.

Volunteered geographic information

Interesting development in the VGI phenomena has been publicly recognized in the last few years for its research opportunities in terms of new usages, effects and practices [1]. Nowadays, VGI initiatives can challenge traditional data suppliers and will grow in the near future [2]. GPS tracking projects [3] and mapping projects as Wikimapia and OSM are more widely used in research and planning. As stated by Craglia et al. [2]: "*Social networking, Web 2.0, and VGI offer also enormous*

opportunities to develop spatial data infrastructures (SDIs) for scientific and policy-support purposes which are yet to be exploited." This paper presents a methodology that handles information integration and space-time analysis and modeling, two needs described by [2]

Network based kernel density estimation

Kernel density estimation (KDE) is a statistical process for spatial smoothing and/or interpolation [4]. It is use in many fields and is widely used by researchers in urban context [e.g. 5,6,7]. KDE is a weighting function *K*, most often a normal function, balancing events accordingly to a distance of influence, the bandwidth. In this paper, KDE gives global views of the event density distribution and show the most interesting bandwidths for later use with NetKDE.
Miller [8] and Batty [9] showed that GIS have to take into account network constraints. Thus, NetKDE uses distances measured along a network rather than Euclidean distances. It has been used to study economic activities spatial distribution in Barcelona [10] and bicycle usage behaviors in Ljubljana [3]. Other similar analyses of density in network constrained environment have been developed [11,12,13]. In this paper, NetKDE produces high resolution views of event density distribution for different bandwidths. These results are later analyzed through time with FMC.

Fuzzy-set map comparison

Urban monitoring is deeply linked to the comparison of maps. Researchers are looking for hot-spots, transition rules and specific patterns between simulation and GIS model maps, and remote sensing images. Recent works present original methods to compare pixel-by-pixel (Kappa statistic) or fuzzy-set (fuzzy Kappa) [14]. This comparison between two maps could deal with nominal, ordinal, interval or ratio scales [14], or even three maps using three-dimensional table [15]. For this paper, NetKDE density class maps are compared to detect spatio-temporal changes. Based on fuzzy-set calculation rules, the FMC compares two maps for their similarities and dissimilarities. This paper uses the Map Comparison Kit (MCK) software from the Research Institute for Knowledge Systems [14]. Empirical tests have been conducted to evaluate the most suitable FMC mathematic function to presents the case study results.

Methodology – Baghdad case study

The case study uses the methodology for better assessment of the spatio-temporal evolution of the density of casualties between 2004 and 2009 in Baghdad area. The Guardian has published a geo-referenced comma-separated values (.csv) file with detailed information from Wikileaks Iraq war logs. OSM network is prepared using topological correction application. Multi-resolution grids have been prepared at two different scale, 200m for global view and 50m for further analysis considering network constrained environment. The Figure 1 presents the data before estimations.

KDE and NetKDE use the same software. The events, the OSM network and the two resolution grid SHP files are imported to PostgreSQL/PostGIS relational database management system (RDMS) using Shape2pgsql application. Spatial objects recovery and other KDE/NetKDE calculations use Python script coupled with Egenix MC base and Psycopg2 API. Visualization of the results is realized with Quantum GIS plugged to the RDMS. Comparison of the result maps is made using the Map Comparison Kit. All the calculations have been completed using normal up-to-date computer using Windows 7.

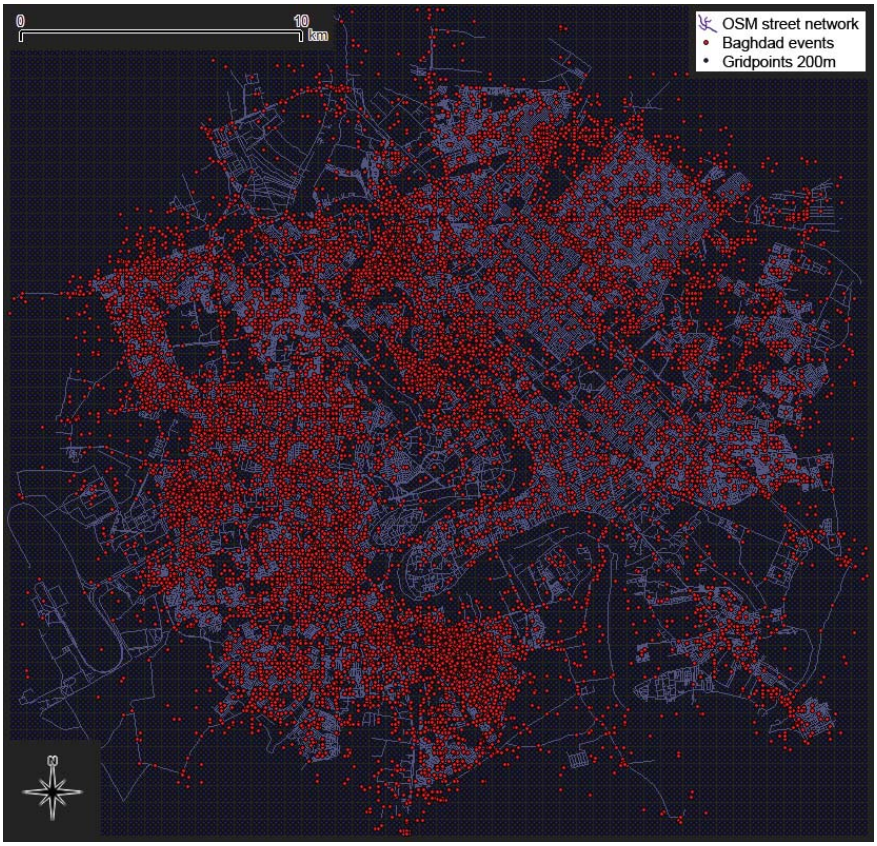

Figure 1: OSM street network, Baghdad events and gridpoints (200m)

Baghdad results

Firstly, the KDE results are presented for low resolution grid (200m). Secondly, the NetKDE results are presented for high resolution grid (50m). Thirdly, the fuzzy-set map comparisons of NetKDE results are presented and discussed between 2004-2005, 2006-2007 and 2008-2009 periods. Discussions of these results look at the further uses and implementations of this methodology. Particular attention is made on urban planning research fields and practices for different type of spatio-temporal data and decision making needs.

References

[1] **Elwood S.** (2008), Volunteered geographic information: key questions, concepts and methods to guide emerging research and practice. GeoJournal, 72, pp. 133-135

[2] **Craglia M. et al.** (2008), Next-Generation Digital Earth – A position paper from the Vespucci Initiative for the Advancement of Geographic Information Science. International Journal of Spatial Data Infrastructure Research, 3, pp.146-167

[3] **Lachance-Bernard et al.** (2011), Network Based Kernel Density Estimation for Cycling Facilities Optimal Location Applied to Ljubljana. International Conference on Computational Science and its Applications – ICCSA 2011 Proceedings, Editors: Murgante B. and Borruso G., Lecture Notes in Computer Science, to be published

[4] **Wang F.** (2006), Quantitative methods and applications in GIS. CRC Press, Boca Raton

[5] **Anselin L. et al.** (2000), Spatial analyses of crime. Criminal Justice, 4, pp. 213-262

[6] **Gatrell A. et al.** (1996), Spatial point pattern analysis and its application in geographical epidemiology. Transactions of the Institute of British geographers, 21, pp. 256-274

[7] **Porta S. et al.** (2009), Street centrality and densities of retails and services in Bologna, Italy. Environment and Planning B: Planning and Design, 36, pp.450-465

[8] **Miller H.J.** (1999), Measuring space-time accessibility benefits within transportation networks: Basic theory and computational procedures. Geographic Analysis, 31, pp.187-212

[9] **Batty M.** (2005), Network Geography: Relations, Interactions, Scaling and Spatial Processes in GIS, Re-presenting GIS, pp. 149-170

[10] **Produit T. et al.** (2010), A Network Based Kernel Density Estimator Applied to Barcelona Economic Activities. International Conference on Computational Science and its Applications – ICCSA 2010 Proceedings, Editors: Taniar D., Gervasi O., Murgante B., Pardede E. and Apduhan B.O., Lecture Notes in Computer Science, 6016(1), pp.32-45

[11] **Borruso, G.** (2008), A GOS Approach for Analyzing Point Patterns in a Network Space. Transactions in GIS, 12, pp.377-402

[12] **Xie Z. and Yan J.** (2008), Kernel Density Estimation of traffic accident in network space. Geography, Geology Faculty Publications

[13] **Okabe A., Satoh T. and Sugihara K.** (2009), A kernel density estimation method for networks, its computational method and a GIS-based tool. International Journal of Geographical Information Science, 23, pp.7-32

[14] **Visser H. and de Nijs T.** (2006), The Map Comparison Kit. Environmental Modelling & Software, 21, pp.346-358

[15] **Pontius R.G., Peethambaram S. and Castella J.-C.** (2011), Comparison of Three Maps at Multiple Resolutions: A Case Study of Land Change Simulation in Cho Don District, Vietnam. Annals of the Association of American Geographers, 101:1, pp.45-62